

Comparison of Six Radioimmunoassay Kits for Human Follitropin in Serum

By L. H. Elvers and J. G. Loeber

National Institute of Public Health, Bilthoven, The Netherlands

(Received August 2, 1982/May 6, 1983)

Summary: Six commercially available radioimmunoassay (RIA) kits for human follitropin (follicle-stimulating hormone; hFSH) were compared with respect to precision, "accuracy", specificity and performance. The intra- and interassay coefficients of variation ranged from 4.7 to 8.3% and from 6.5 to 13.3%, respectively. The recovery of added follitropin ranged from 88 to 176%. The most important conclusion of this study is that an accurate calibration of the kit standard by the manufacturer is essential for minimizing the variation between kits.

Vergleich von sechs Radioimmunoassay-Bestecks zur Bestimmung von menschlichem Follitropin im Serum

Zusammenfassung: Sechs kommerziell erhältliche Radioimmunassay-Bestecks zur Bestimmung menschlichen Follitropin wurden hinsichtlich Präzision, Richtigkeit, Spezifität und Durchführung verglichen. Die Variationskoeffizienten betrugen in der Serie 4,7 bis 8,3%, von Tag zu Tag 6,5 bis 13,3%. Die Wiederfindung zugesetzten Follitropins reichte von 88 bis 176%. Die wichtigste Schlußfolgerung aus dieser Untersuchung ist, daß eine richtige Kalibrierung des Standards im Testbesteck durch den Hersteller für die Verminderung der Variation der mit den Bestecks erzielten Ergebnisse von ausschlaggebender Bedeutung ist.

Introduction

Standardization, quality control, reference sera have become common concepts in clinical laboratories. Besides using internal quality control schemes most laboratories participate in one or more external quality control programmes. At regular intervals the participants receive samples for analysis. The returned results are evaluated and each participant is informed about his result in comparison with all other participants or with the users of the same reagents. The reports of such external quality control programmes show that the use of different methods and reagents adds greatly to the between-laboratory variation. The present study, on six kits for the determination of the pituitary hormone follitropin, was undertaken to determine how these kits differ and how they may be improved.

¹⁾ Reference to a company and/or product is only for the purpose of information and identification and does not imply approval or recommendation of the company and/or product by the National Institute of Public Health to the exclusion of others.

Materials and Methods

In this study, carried out in the first trimester of 1982, the following kits¹⁾ were used and further referred to as mentioned in parentheses:

1. RIA-gnost® hFSH; Behringwerke AG, D-3550 Marburg, Germany (Behring); batch Nos. W1007, W1010 and A1002.
2. Biodata FSH-DAB; Hypolab S.A., CH-1267 Coinsins, Switzerland (Biodata); batch Nos. 103211 and 103213.
3. RIA-mat® FSH; Byk-Mallinckrodt, D-6057 Dietzenbach-Steinberg, Germany (Byk); batch Nos. 2021, 2041 and 2081.
4. DPC FSH-RIA; Diagnostic Products Corporation, Los Angeles, California, U.S.A. (DPC); batch Nos. 052, 056 and 059.
5. IRE-FSH-RIA; IRE, B-6220 Fleurus, Belgium (IRE); batch Nos. 181181, 1811281 and 180182.
6. FSH-RIA IM72; The Radiochemical Centre, Amersham, U.K. (RCA); batch Nos. 183, 187 and 190A.

The following International Reference Preparations (IRP) were kindly donated by the National Institute for Biological Standards and Control (London, U.K.) and are further referred to as mentioned in parentheses: the 1st and 2nd IRP for human follitropin (FSH)/lutropin (LH) for bioassay (69/104 and 78/549, respectively), the 1st IRP for human thyrotropin (TSH) for immunoassay (68/38), the 1st IRP for human lutropin (LH) for immunoassay (68/40), and the 1st IRP for human chorionic gonadotropin (CG) for immunoassay (75/537).

Bovine serum albumin (BSA; Poviet) was purchased from Organon, Oss. Serum and plasma samples with various follitropin concentrations, were kindly donated by Prof. dr. J. H. H. Thijssen and Dr. G. J. P. Alsbach, Academic Hospital, Utrecht and Dr. W. Schopman, Bergweg Hospital, Rotterdam.

All experiments were carried out by the same technician. The RIA results were calculated using a computer program developed by Rodbard (1).

Results and Discussion

Technical Data

Table 1 shows certain technical data for each kit, such as the reference preparation used for calibration of the kit standard, conditions for (pre-)incubation and the separation method.

The World Health Organization (WHO) has recently established International Reference Preparations for the immunoassay (2) of a number of hormones, but not follitropin. In the meantime, WHO recommends that 69/104, which has been established for bioassay, should also be used for the follitropin immunoassay. Recently, this preparation has been replaced by 78/549 which is equipotent (3). Two kits, DPC and RCA, still use the obsolete urinary reference preparation 2nd IRP of Human Menopausal Gonadotropin (hMG), but provide a conversion factor for calculation in terms of 69/104 and 78/549. In one experiment with each kit a dilution series of 69/104 was compared with one of 78/549. In all cases the two curves were coincident. Hence, for brevity, all further results are expressed in terms of 78/549.

The amount of radioactivity per 100 tubes for the kits investigated here varies from 44 kBq (Biodata) to 148 kBq (Byk). This last figure appeared to correspond to 40000–50000 counts per minute per tube. It is obvious that both the manufacturer and the user are interested in counting times that are as short as possible, but with sufficient counting precision. However, from the point of view of environmental protection as well as radiation exposure of the technician it is important to keep the amount of radioactivity as low as possible. To us, these two arguments are more important than the relatively small increase in counting time.

Precision

For the determination of the intra- and interassay variation, eight sera, coded A to H, were analysed in triplicate with each kit. In each experiment a dilution series of 78/549 was assayed as well. This preparation was dissolved and diluted in phosphate-buffered saline pH 7.4 containing 10 g/l bovine serum albumin for analysis by Biodata, IRE and RCA, and in the corresponding matrix of the kit standard for analysis by Behring, Byk and DPC.

For radioimmunoassays the intra- and interassay coefficients of variation are dependent on the analyte concentration. Therefore, for each serum sample the results of the six experiments per kit were submitted to an analysis of variance according to *McDonagh et al.* (4), thus yielding an estimate both for the cumulative intra-assay variation and the cu-

Tab. 1. Technical data

	Standard		Radio-activity (kBq) ²⁾	Pipetting steps	Separation method	Incubation time (h)		Separation B/F (min)	
	IRP	matrix				without tracer	with tracer	incubation	centrifuge
Behring	69/104	HS ¹⁾	74	5 ³⁾	immuno-sorbent	—	16–24	15	2 × 20
Biodata	69/104	B	44	5	DA	—	3 16–24 ⁴⁾	60 120	30
Byk	69/104	HS	148	4	DA + PEG	— 16–24	16–24 ⁴⁾ 16–24	30	30
DPC	2 nd IRP hMG	PS	74	4	DA + PEG	2/37 °C 2/37 °C	2/37 °C 15 ⁴⁾	5	20
IRE	69/104	B	74	6 ³⁾	immuno-sorbent	—	18–24	60	2 × 10
RCA	2 nd IRP hMG	B	92	5	DA + (NH ₄) ₂ SO ₄	—	16–24	5	30

¹⁾ Abbreviations: HS: human serum; B: buffer; PS: protein-containing solution; PEG: polyethyleneglycol; DA: double antibody

²⁾ number of kBq per 100 tubes (1 kBq = 0.027 µCi)

³⁾ including washing of the precipitate

⁴⁾ incubation times chosen in the study

mulative interassay variation. These estimates are shown in tables 2 and 3, respectively and reflect the so called „precision profile“ (5). To facilitate the overview, the median of the estimates per kit is also given. Behring shows the highest intraassay variation and DPC the highest interassay variation, but the differences with the other kits are relatively small.

Accuracy

The above experiments provided the mean follitropin concentration of each sample for each assay kit, and an overall mean for each sample based on analysis by all six kits. Table 4 shows the results. To determine whether any kit systematically yields high or low follitropin values the kits have been ranked from 1 to 6 according to their results for each serum sam-

ple. The kit yielding the lowest follitropin value for serum A was given number 1, the kit yielding the next higher value was given number 2, etc. The same procedure was followed for the other sera and the results per kit were summed.

According to *Thompson & Willke* (6) for 8 sera and 6 kits the 5% significance values for the sum of the ranking numbers are 15 and 41. The ranking numbers and the sums are presented in table 5. It appears that Byk measures systematically lower and Behring higher than the other kits. Since in all experiments a dilution series of 78/549 was assayed it was possible to read the serum samples on this curve (tab. 6) and to recalculate the ranking (tab. 7). It is striking that Behring then shows a "normal" level whereas Byk still measures relatively low and IRE

Tab. 2. Intra assay variation for the samples A to H calculated from the results in triplicate of 6 assays.

	Coefficient of variation (%)								Median
	A	B	C	D	E	F	G	H	
Behring	16.2	13.7	9.9	9.9	6.7	3.3	6.3	4.7	8.3
Biodata	10.7	6.0	4.5	6.9	5.6	10.9	3.2	6.0	5.8
Byk	6.8	10.8	3.1	3.3	3.7	5.6	2.1	8.0	4.7
DPC	18.9	13.8	7.4	7.5	5.1	6.2	3.6	5.2	6.8
IRE	8.2	5.1	6.2	5.8	5.9	6.4	4.7	4.8	5.9
RCA	6.0	7.8	6.4	5.8	3.8	5.6	4.5	4.1	5.6

Tab. 3. Inter assay variation for the samples A to H calculated from the results in triplicate of 6 assays.

	Coefficient of variation (%)								Median
	A	B	C	D	E	F	G	H	
Behring	16.2	26.7	12.6	10.4	11.2	6.5	12.2	6.2	11.8
Biodata	32.8	9.4	6.2	10.1	6.4	12.0	7.4	6.2	8.4
Byk	17.3	12.8	5.0	6.4	4.1	6.7	5.7	8.0	6.6
DPC	25.1	27.1	13.0	15.6	13.5	10.4	5.7	7.0	13.3
IRE	14.7	9.0	9.7	10.4	9.7	9.3	8.0	7.9	9.5
RCA	6.9	27.0	9.5	6.5	5.5	6.4	4.9	5.1	6.5

Tab. 4. Mean follitropin concentration of the samples A to H with respect to the kit standard calculated from 6 assays.

	Follitropin concentration (IU/l)							
	A	B	C	D	E	F	G	H
Behring	3.6	4.6	7.0	11.1	16.4	37.7	43.2	68.3
Biodata	1.2	2.2	3.7	6.1	8.8	19.5	27.0	37.1
Byk	1.2	1.2	3.5	6.0	9.1	17.9	23.9	34.1
DPC	1.9	3.1	4.4	6.6	8.8	18.8	28.9	38.2
IRE	2.4	4.4	6.7	10.9	15.2	35.0	53.2	73.9
RCA	2.7	2.2	3.7	5.9	8.4	18.2	23.3	36.2
\bar{x}	2.2	3.0	4.8	7.8	11.1	24.5	33.2	48.0
CV (%)	43	46	33	32	33	38	37	38

Tab. 5. Ranking numbers for the samples A to H calculated with respect to the kit standard

	A	B	C	D	E	F	G	H	Sum ¹⁾
Behring	6	6	6	6	6	6	5	5	46
Biodata	1.5	2.5	2.5	3	2.5	4	3	3	22
Byk	1.5	1	1	2	4	1	2	1	13.5
DPC	3	4	4	4	2.5	3	4	4	28.5
IRE	4	5	5	5	5	5	6	6	41
RCA	5	2.5	2.5	1	1	2	1	2	17

¹⁾ For 6 kits and 8 samples the 5% significance values are 15 and 41 (ref. (6))

Tab. 6. Mean follitropin concentration for the samples A to H with respect to follitropin 78/549.

	Follitropin concentration (IU/l)							
	A	B	C	D	E	F	G	H
Behring	2.2	2.8	4.4	6.9	10.2	23.9	27.4	43.7
Biodata	1.4	2.9	4.8	7.8	11.0	23.4	32.1	44.0
Byk	0.8	0.8	2.7	5.2	8.1	17.0	23.9	36.8
DPC	1.7	2.8	4.1	6.2	8.6	18.9	29.4	39.2
IRE	2.1	3.7	5.7	9.1	12.7	29.3	44.7	62.4
RCA	2.4	1.9	3.4	5.5	8.1	18.6	24.1	38.8
\bar{x}	1.8	2.5	4.2	6.8	9.8	21.8	30.3	44.2
CV (%)	34	40	25	22	19	21	26	21

Tab. 7. Ranking numbers for the samples A to H calculated with respect to follitropin 78/549.

	A	B	C	D	E	F	G	H	Sum ¹⁾
Behring	5	3.5	4	4	4	5	3	4	32.5
Biodata	2	5	5	5	5	4	5	5	36
Byk	1	1	1	1	1.5	1	1	1	8.5
DPC	3	3.5	3	3	3	3	4	3	25.5
IRE	4	6	6	6	6	6	6	6	46
RCA	6	2	2	2	1.5	2	2	2	19.5

¹⁾ For 6 kits and 8 samples the 5% significance values are 15 and 41 (ref. (6))

relatively high. For follitropin there is no definitive method so it cannot be stated that the results of any particular kit are wrong. In addition, the ranking depends on the choice of kits being investigated. Nonetheless, it was decided to further investigate the calibration of the kit standards.

Calibration of kit standards and recovery experiments

By direct comparison of the calibration curves of the kit standard and 78/549 a potency was calculated for each kit. The ratio of this potency and the one stated by the manufacturer is shown in table 8. In addition, a recovery experiment was carried out as follows. To a plasma sample with a low follitropin concentration, 78/549 was added in amounts of 2.5, 5, 10, 20 and

40 IU per litre. For each kit the spiked samples were assayed in one experiment in quadruplicate. The follitropin concentration was calculated with respect to the kit standard and to 78/549. The percentage recovery is shown in table 9. From table 8 it is clear that the potency of the kit standards with respect to 78/549 differ from the expected value of 1.0, mainly for Behring and Biodata and to a smaller degree for Byk and IRE. The kit standard of Biodata has been underestimated whereas the others have been overestimated. These discrepancies are not caused by matrix effects, since for each kit 78/549 was dissolved in the matrix prescribed for the kit standard. For Behring, Biodata and IRE these findings do agree with the results of the recovery experiment (tab. 9); the recovery with respect to 78/549 was

Tab. 8. The ratio of the immunochemical potency of the kit standard with respect to 78/549 as observed in this study and as stated by the manufacturer.

Behring	0.68	(0.64–0.73) ¹⁾	n = 2 ²
Biodata	1.27	(1.24–1.31)	n = 4
Byk	0.86	(0.82–0.91)	n = 2
DPC	0.96	(0.92–1.00)	n = 3
IRE	0.85	(0.83–0.86)	n = 6
RCA	0.97	(0.94–0.99)	n = 6

¹⁾ 95% confidence limits

²⁾ number of assays

Tab. 9. Percentage recovery of follitropin (FSH) 78/549 added to plasma, calculated with respect to the kit standard and to follitropin (FSH) 78/549. Additions were 2.5, 5, 10, 20 and 40 IU follitropin (FSH) 78/549 per liter plasma.

	Recovery (%), $\bar{x} \pm s$, n = 20	
	Kit standard	FSH 78/549
Behring	176 \pm 22	114 \pm 15
Biodata	88 \pm 10	96 \pm 11
Byk	105 \pm 14	106 \pm 8
DPC	109 \pm 8	106 \pm 8
IRE	132 \pm 7	109 \pm 6
RCA	101 \pm 9	108 \pm 8

normal whereas the recovery with respect to the kit standard was too high for Behring and IRE and too low for Biodata.

It is obvious that incorrect calibration of the kit standard adds to the variation between kits. This is clearly seen when tables 4 and 6 are compared. The median coefficient of variation of the overall mean follitropin concentration per sample decreases from 38%, when the follitropin concentration is calculated using the kit standard, to 23% when the follitropin concentration is calculated using 78/549. It follows that a more precise calibration of the kit standard will mean a great improvement in the process of standardization and quality control.

Specificity

Similarities in the structures of follitropin, thyrotropin, lutropin and chorionic gonadotropin sometimes lead to cross reactions in assay systems. The specificity of the kits was tested with the preparations TSH 68/38, LH 68/40 and hCG 75/537 in amounts up to 100 mIU, 400 IU and 400 IU per litre serum, respectively. Care was taken to dissolve these in the same matrix as the respective kit standard.

Usually, specificity is expressed as the percentage cross reaction, i. e. the mass of the analyte causing 50% decrease of initial binding divided by the mass of the cross reacting compound causing a similar decrease. In the present study, such a calculation is impossible for two reasons. Firstly, for 5 of the 6 kits to reach a 50% decrease of the initial binding a tremendous amount of LH 68/40 and hCG 75/537 would have to be used; secondly, it is meaningless to calculate a ratio of e.g. International Units of follitropin and International Units of lutropin. Hence, the specificity is expressed as the concentration of thyrotropin, lutropin or chorionic gonadotropin that causes a decrease of the binding to 90% of the initial value (tab. 10). In general, the TSH 68/38 curves ran parallel to 78/549. Moreover, the relative potency was approximately the same for all kits. It is concluded, therefore, that the thyrotropin „cross reaction“ in fact is caused by follitropin contamination in TSH 68/38. The lutropin cross reaction is too low to cause interference with the measurement of follitropin in physiological samples, except for RCA. The cross reaction of human chorionic gonadotropin (hCG) 75/537 in the case of Behring, Byk and in particular RCA is rather high. To show this, two sera of pregnant women (hCG > 10000 IU/l) were assayed. It can be seen from table 11 that these 3 kits give relatively high values. Concerning specificity it must be stated that only Biodata and RCA provide figures. Behring does not mention specificity at all

Tab. 10. Specificity of the antisera expressed as the concentration of human follitropin (hFSH), human thyrotropin (hTSH), human lutropin (hLH) and human chorionic gonadotropin (hCG), respectively, causing a decrease of the binding to 90% of the initial binding.

	hFSH 78/549 IU/l serum	hTSH 68/38 IU/l serum	hLH 68/40 IU/l serum	hCG 75/537 IU/l serum
Behring	1.3	29	160	400
Biodata	1.1	28	>400	>400
Byk	0.4	36	240	240
DPC	1.3	38	>400	>400
IRE	0.9	20	360	>400
RCA	0.6	10	12	14

Tab. 11. Follitropin concentration (IU/l) in two sera from pregnant women (hCG > 10000 IU/l serum).

	Serum I	Serum II
Behring	3.5 ¹⁾	5.2
Biodata	1.4	1.7
Byk	3.0	3.8
DPC	1.4	1.4
IRE	1.9	1.9
RCA	4.0	5.1

¹⁾ mean of at least 2 assays in triplicate

and, like Byk, does not warn against cross reaction with chorionic gonatropin. IRE and DPC do provide cross reaction figures for lutropin, thyrotropin and chorionic gonadotropin but this information is worthless since the purity of the preparations used is not stated. We realise that our choice of the thyrotropin, lutropin and chorionic gonadotropin preparations for the study of the specificity can be debated with respect to their purity. On the other hand these international reference preparations have been very well documented and are freely available.

References

1. Rodbard, D. & Lewald, J. E. (1970) *Acta Endocrinol. (Kbh.)* Suppl. 147, 79–103.
2. World Health Organization (1975) WHO Expert Committee on Biological Standardization, 26th Report. Technical Report Series nr 565.
3. World Health Organization (1981) WHO Expert Committee on Biological Standardization, 31th Report. Technical Report Series nr 658.
4. McDonagh, B. F., Munson, P. J. & Rodbard, D. (1977) *Comp. Prog. Biomed.* 7, 179–190.
5. Ekins, R. P. (1978) In: *Radioimmunoassay and related procedures in medicine*, Vol. II, 6–20, IAEA, Vienna.
6. Thompson, W. A. & Willke, T. A. (1963) *Biometrika* 50, 375–380.

J. G. Loeber
Laboratory for Endocrinology
National Institute of Public Health
P.O. Box 1
NL-3720 BA Bilthoven